

Factors influencing the choice of relatum in referring expressions generation: animacy vs. position

Adriana Baltaretu (a.a.baltaretu@tilburguniversity.edu)

Emiel Krahmer (e.j.krahmer@tilburguniversity.edu)

Alfons Maes (maes@tilburguniversity.edu)

Tilburg Center for Cognition and Communication (TiCC), School of Humanities, Tilburg University,
PO Box 90153, 5000 LE, Tilburg, The Netherlands

Abstract

One way of identifying objects using spatial language is by relating a target (“the ball”) to another entity, the relatum (“the man”) via relational descriptions (“the ball in front of the man; the ball in between the man and the closet”). In a production experiment, we investigated two factors that might influence the choice of relatum: animacy and visual position. Results show that animacy did not influence the relatum preferences. However the position of the entity was found to affect to a great extent participants' choice, with left entities being more likely to be chosen as (first) relatum than right ones.

Keywords: relational descriptions; animacy; spatial position; relatum choice; referring expressions.

While communicating, speakers refer to objects in a manner that selects some aspects of the visual experience and ignores others. For example, to identify an object via an exophoric reference, the speaker can refer to perceptual (e.g., colour) or spatial (“the ball in front of the man”) features. Expressing spatial relations indicates the location of one object (the target) in relation to another object (the relatum; Levelt, 1989). Indicating the location of an object is an aspect of identification (Miller & Johnson-Laird, 1976) and an ordinary daily life situation.

Identification using relational descriptions is a particular bottom-up (stimulus-driven) type of reference. It becomes possible to identify an object via its location provided that the visual scene is coherent and affords such a description and referring to the object's location facilitates identification. Relational descriptions are understudied in the field of referential communication. On the one hand, previous experiments on referential communication control tasks in such a way that spatial location information cannot be used as a potential disambiguating feature. Many psycholinguistic experiments forbid the use of this feature and most algorithms that generate referring expressions (RE) fail to produce relational descriptions (Krahmer & van Deemter, 2012). On the other hand, compared to other attributes (such as colour and size), relational descriptions are a dispreferred identification feature in human communication. It is generally considered more cognitively expensive to describe two objects (a target and a relatum)

than one (principle of least effort, e.g., Zipf, 1949). People prefer to describe an object by referring to simple properties, and use relations only when other properties do not suffice (Krahmer & Theune, 2002). Thus, various attempts to extend existing algorithms by allowing relational descriptions (Krahmer & Theune 2002; Kelleher & Kruijff, 2006) assume that relational properties are the least preferred attributes.

However, recent studies have shown that humans do use RE containing relations, even when unnecessary (Viethen & Dale, 2008). Viethen and Dale's (2008) results suggest that in simple scenes, where objects can be distinguished without using location, participants produce relational descriptions in about one third of the trials. It is not clear when relations are preferred and there are quite a few situations in which location is the most likely identification feature. For example, rather than producing a long expression that includes three attributes, but no location, it might be the case that using one relation is enough to identify the target (e.g., “give me the pen by the book”). Also, the use of locational descriptions can be beneficial for the addressee (Arts, Maes, Noordman, & Jansen, 2011). Far from being a trivial feature, location is employed in a meaningful way in different forms of descriptions and visualizations. People naturally refer to location in a variety of situations (landmark identification, interaction with conversational agents, visual communication: maps, graphs, charts) and fields of work (architecture, design, geosciences, engineering) (for a review see Tversky, 2011). Identifying via location is not an unusual act due to the fact that space itself is a pervasive dimension of everyday life. We make use of space in gestures and discourse (“the point *here* is”, “a topic is *central* to a debate”, “somebody has *fallen into* a depression”, Lakoff & Johnson, 1980), in actions (lining up the ingredients for a recipe in order of use, Kirsh, 1995) and in reasoning (by mapping time onto space, etc).

Thus, when relations can be used in identification scenarios, what defines the choice of relatum? Although a large body of research in the field of spatial cognition is devoted to the usage of spatial prepositions (for a review see Coventry & Garrod, 2004), most of those studies are focused on the problem of localization rather than identification. In other words, studies of spatial language production have generally limited tasks so that speakers are

forced to refer to already agreed upon objects, based solely on their spatial locations. On the other hand, identification is a matter of choice: the main goal of a referring expression is to identify a target object out of several candidates, in this specific case by making use of the spatial position. In everyday situations there are often many objects in the proximity of the target, each of which could be a potential relatum. Yet, different features can make an object more salient, thus more likely to be chosen and mentioned as relatum (Beun & Cremer, 1989; Miller, Carlson & Hill, 2011). Objects can be salient on different dimensions. The saliency of an object can be influenced both by bottom-up (e.g., spatial, perceptual, functional-interactive features; Miller, Carlson & Hill, 2011; Kelleher & Kruijff, 2006) and top-down factors (intrinsic characteristics of relata, task, prior knowledge; Gapp, 1995).

Little is known on how these factors influence relatum choice. The scarcity of research on this topic is a major stumbling block for developing algorithms that can generate relational descriptions. In this paper, we study the trade-off between two factors relevant for relatum choice: animacy and visual position. Animacy is an inherent characteristic of the entities referred to, while visual position is an extrinsic one.

There are reasons to believe that animacy may affect the production of spatial relational descriptions. Animacy is an inherent characteristic of the entities referred to, and the categorization of objects into animate/inanimate develops early in the childhood (Mandler, 1992). First, animacy is considered a salient feature of noun phrases (e.g. Clark & Begun, 1971) and it was shown to influence discourse salience. Animate entities are conceptually highly accessible, retrieved and processed more easily, influencing language production (e.g., Prat-Sala & Branigan, 2000). Thus animate entities are more likely to be assigned to subject position and inanimate entities as objects. Second, animacy influences visual attention. Some of the earliest studies showed that human subjects look preferentially at people and faces (Yarbus, 1967).

In relation to spatial cognition, animacy has been reported to influence relational descriptions, by triggering spontaneous perspective taking (Tversky & Hard, 2009), influence preposition choice (Feist & Gentner, 2003) and performance in spatial tasks (Böckler, Knoblic & Sebanz, 2011). In referring expressions production, the choice of referring expression is assumed to be influenced by the referent's saliency, thus, animacy may influence which entity a speaker chooses to mention first. In the current study we manipulated animacy in the visual modality.

However, the choice for relatum might be influenced by another (extrinsic) factor which is present in any visual scene: the item's (visual) position. Different studies suggest that the spatial routines employed by reading and writing can have an impact on spatial organization, memory, and visual attention. For instance, the directionality of the language system influences the location where speakers tend to look first and remember information (Chan & Bergen,

2005). This implies that, at least in western cultures, people "read" visual scenes from left to right. This bias is shown from a very young age (Tversky, Kugelmass, & Winter, 1991) in graphical representations of spatial and temporal relations.

The directionality of the language system not only affects cognitive non-linguistic processes but also linguistic ones. The scanning habit influences spatial representations of objects and as a result, speakers with different writing systems show different patterns of sentence production. For example, Maas & Russo (2003) observed that in a sentence-picture matching task involving motion events, Italians (whose writing system runs from left to right) tend to place the agent on the left of the patient, while Arabic speaking participants (whose system goes from right to left) place the agent on the right of the patient. The same pattern of results is found with other language pairs and in a task that does not include spatial representation of actions (Chan & Bergen, 2005).

The left-right bias was also observed in a clinical population (Chatterjee, 2001). Subjects suffering from agrammatism, an aphasic syndrome, showed a left-right bias both in language production (in describing visual scenes) and comprehension (in matching sentences with pictures). In addition, congruent evidence comes from studies in the psychology of art showing that reading habits influence visual preferences (subjects prefer the pictures possessing the same directionality as their reading system) and influence the way we direct our attention (Chokron & De Agostini, 2000). Explanations for this bias are given in terms of iconic scanning routines of viewers based on their writing directions.

In the light of these studies it seems there might be a trade-off between two factors: (1) people prefer animate relata over non-animate ones, and (2) people "scan" scenes from left to right, thus the item in the picture's left area will be chosen or named first. To study the factors influencing the relatum choice, a production experiment was carried out: participants were asked to identify a target object in between two relatum objects: one animate, one inanimate and one left, one right. Both relatum candidates were present in the scene, while their position was systematically changed across the scenes.

Method

Participants

64 Dutch-speaking students of Tilburg University (50 women, mean age 22) were given partial course credits for the participation in the experiment. Due to technical problems, speech data of 4 participants was excluded from the analysis.

Stimuli and Design

The experiment consisted of 96 greyscale trial scenes (24 critical trials, 72 filler trials). The 24 critical trial scenes included a target item (a ball), and two relatum candidates

(an inanimate and an inanimate one) arranged in a particular (left/right) position. The scene also included a distractor (an identical ball) in order to prevent an easy identification strategy using type only. The inanimate items were everyday objects, easily identifiable, with a clear front/back axis and of roughly equal size as the animate items (see figure 1). The target was placed in front of the two potential relata, located at equal distance from both. The angle between the left relatum, the target and the right relatum is 90°. The viewer's perspective is aligned with the target in such a manner that the relata are symmetrically positioned at the left/right of the target (see figure 2).

The relata items (8 animate and 8 inanimate) were pretested: ten subjects were presented with pictures similar to those used in the critical trials and had to name the inanimate items, the persons' gender and profession.

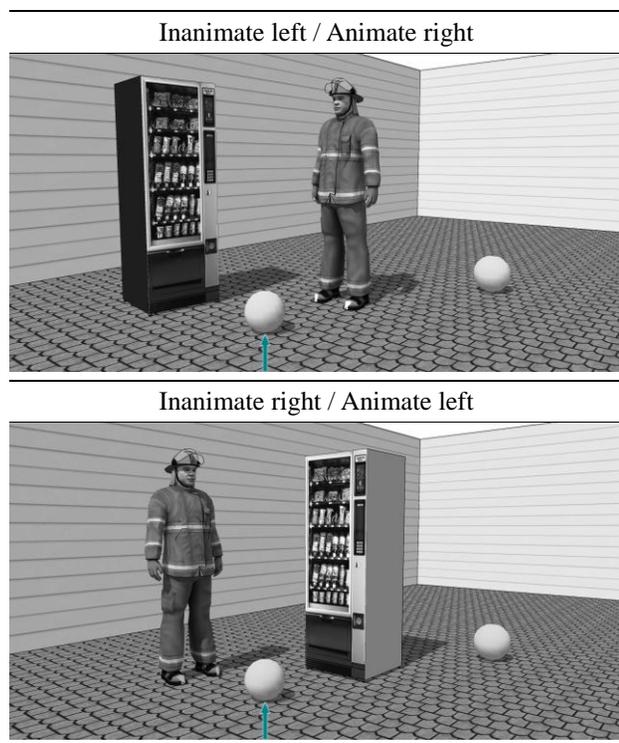


Figure 1: Example of critical trials.

An inanimate item was included in the critical trial if: (1) it was referred to with the same noun in minimum 50% of the cases, and (2) if the other nouns used to refer to it are compounds nouns such as in “kast” – “ladenkast”. An animate item was chosen if: (1) the character's gender was recognized in all cases and (2) if the character's profession was recognized in 80% of the cases. From 64 possible combinations of objects with people, 24 random couples were chosen. Each object and each person was presented 3 times in a different combination.

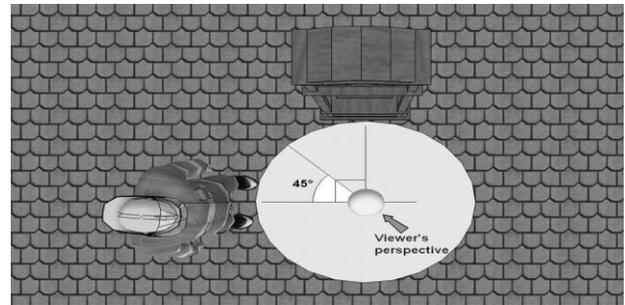


Figure 2: Items' positions.

The relata's position (left/right of the screen) was counterbalanced (see figure 1) and half of the critical trials were presented in mirror image¹. Thus, the distance between the distractor ball and the relatum candidates was counterbalanced. In addition, each critical trial was followed by 3 fillers to prevent a training effect. The fillers required identification via different identification strategies (such as type and size). The pictures were created using Google SketchUp 8 (3D Warehouse library), and the stimuli were presented electronically using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).

Procedure

At the beginning of the experiment the instructions were shown in written form on the monitor, informing the participants that they had to refer to an object marked with an arrow. Participants were instructed to refer to it in such a way that the next participant (a fictitious addressee) could draw the arrows on a new set of identical pictures.

Descriptions using a deictic perspective or absolute location (e.g., “the ball on the left”) were discouraged, by telling the speaker that the addressee would receive the same image, but it might be (or might not be) a mirror version. The picture remained on the screen until the participant responded and pressed a button to continue. The experiment started with 3 practice trials followed by 96 experimental trials.

Participants were unaware of the goals of the experiment and of the type of trials they were attending. No feedback was given during the task. The experiment lasted approximately 15 minutes.

Results

For the purpose of the current analysis the descriptions of the critical trials were transcribed and annotated. Participants were found to use one of two possible description strategies; either selecting a single relatum (e.g., “the ball *in front of* the man”) or mentioning both (e.g., “the ball *in between* the drawer and the man”). All participants consistently used one of the two identification strategies: 48

¹ The full set of critical trial pictures used in the experiment can be seen at <https://sites.google.com/site/spatialreference/stimuli>

named both relatum candidates, and 12 chose one relatum. Both strategies are considered valid for identification via a relatum. In the case of the *in between strategy*, the order in which entities were mentioned is considered to witness the importance of animacy or visual position. These two groups were analysed separately using a Pearson chi-square test.

For the *in between strategy*, there was a significant association between the item's visual position and its mention in the first slot after the preposition ($\chi^2(1) = 56.82, p < .001$). Figure 3 reveals that in 61% of the descriptions, the left-most relatum was mentioned first, irrespective of its animacy.

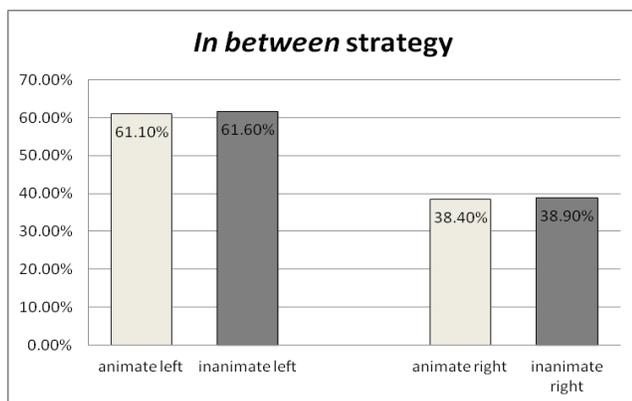


Figure 3: Percentages of descriptions mentioning an (in)animate relatum first as a function of left-right position produced for the *in between strategy*.

The data in the *in front of strategy* presents the same pattern ($\chi^2(1) = 7.818, p = 0.005$). The percentage of descriptions choosing the left-most relatum, irrespective of its animacy, is shown in Figure 4.

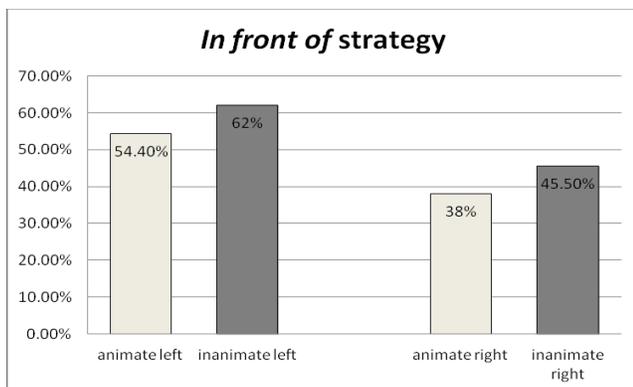


Figure 4: Percentages of descriptions mentioning an (in)animate relatum first as a function of left-right position produced for the *in front of strategy*.

Discussion and Conclusion

The results show that when people identify objects via a relatum, whether the object is animate or not does not seem to influence their choice. However the position of the entity

was found to affect to a greater extent participants' choice, with left entities being more likely to be chosen as (first) relatum than right ones. Participants used two strategies to describe the scene: they either mentioned both relatum candidates or selected one of them. In addition, once a strategy had been chosen, speakers used it consistently; no one changed strategy during the course of describing the 24 critical trials. The items' visual position was the only statistically significant factor and the effect observed was replicated across the two description strategies.

The left/right bias, possibly caused by the directionality of the language system, has been observed to influence different aspects of spatial cognition and language production. In a next study, we will use the same design for testing the visual position's effect on relatum selection with languages that have different writing systems such as Arabic and Iranian.

Within this particular spatial layout, animacy did not influence the choice of relatum. Animate entities were chosen as relatum almost equally often as the inanimate ones. This result is not affected by word frequency: 90% of the participants referred to the animate entity using highly frequent words such as “de vrouw / de man” (the woman / the man). The results are intriguing knowing that animacy is considered an influential factor in language production. A possible explanation may be found in studies that have shown that preferred relata (or “landmarks”) are objects with a stationary setting within a certain reference frame (Talmy, 1983). One corpus study suggests that relatum objects tend to be inanimate and stable (de Vega, Rodrigo, Ato, Dehn, & Barquero, 2002), thus animacy might not be a preferred feature in choosing a relatum. However, in our study inanimate entities are as (dis)preferred as animate ones. This equality might be due to two competing forces: animacy and landmark stability; however this hypothesis should be tested in a further study.

As for the strategies used, approximately a quarter of the subjects chose one relatum, thus producing the X *in front of* Y description. The chance of picking one of the relata candidates was not influenced by the position of the distractor (the further away the relatum is from the distractor ball, the less ambiguous). The location of the distractor was counterbalanced regarding both its position on the screen and its relative position with respect to the relatum candidates. Participants didn't seem to prefer the relatum which was furthest away from the distractor object, but rather the relatum located on the left side of the screen. Moreover, during the debriefing, participants did not mention using this selection criterion.

Most of the participants identified the target using the preposition “tussen” (Dutch for “between”). This preposition describes the location of the target in relation with both relata. This option can be explained by the entities' position and orientation in the spatial layout. The “between” preposition assumes a specific (almost linear) ordering of entities (van der Zee & Watson, 2005). However, compared with other locative prepositions

“between” is a cognitively more expensive one. Compare the answers: “the ball is in front of the closet” vs. “the ball is between the closet and the man”. The former answer seems easier, because it involves relating the target to only one referent and is shorter. Still the vast majority of the participants opted to identify the object by relating it to both relatum candidates, starting from the left side. Is this preposition more accurate? Mentioning one relatum only is only partly accurate: the target is in front of both relatum. Insofar, the relation between the preposition and the specific situation in which is produced has not been thoroughly investigated. There is one characteristic of the visual scene which might have encouraged participants to choose this preposition. The relatum candidates are orientated symmetrically relative to the target, but they are not aligned with the participant view point (oblique perspective; see figure 2). Even though the target falls directly on the relata front axis, the viewer’s perspective seems to distort the spatial templates, thus the *in front of* strategy becomes less favoured. Kelleher, Ross, Mac Namee, & Sloan(2010) found that presenting a landmark from an oblique point of view relative to the addressee, produced anomalies in what was generally accepted to be the geometrically “in front of” area of a relatum. The “in front of” template was distorted to a large extent. Unfortunately, effects of oblique perspective are not discussed in the spatial language literature.

Finally, spatial descriptions of targets often involve the selection of a reference object out of several candidates. Apart from visual saliency, other aspects can influence the selection (Gapp, 1995; Miller, Carlson & Hill, 2011). This experiment confronted two factors when two candidates are present in the visual scene: an extrinsic factor (visual position) and an intrinsic one (animacy). The results suggest that referring expressions algorithms should take into account the visual position of objects in the scenes when selecting a relatum.

References

- Arts, A., Maes, A., Noordman, L. G. M., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, 43, 361-374.
- Beun, R. J., & Cremers, A. (1998). Object reference in a shared domain of conversation. *Pragmatics and Cognition*, 6, 121-152.
- Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: Effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, 211, 531-545.
- Chan, T., & Bergen, B. (2005). Writing Direction Influences Spatial Cognition. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, (pp. 412-417). Mahwah, NJ: Erlbaum.
- Chatterjee, A. (2001). Language and space. Some interactions, *Trends in Cognitive Sciences*, 5, 55–61.
- Chokron, S., & De Agostini, M. (2000). Reading habits influence aesthetic preference. *Cognitive Brain Research*, 10, 45–49.
- Clark, H. H., & Begun, J.,S. (1971). The semantics of sentence subjects. *Language and Speech*, 14, 36-46.
- Coventry, K. R. & Garrod, S. C. (Eds). (2004). *Saying, Seeing and Acting: the psychological semantics of spatial prepositions*. USA: Psychology press
- de Vega, M., Rodrigo, M. J., Ato, M., Dehn, D. M., & Barquero, B. (2002). How nouns and prepositions fit together: An exploration of the semantics of locative sentences. *Discourse Processes*, 34, 117–143.
- Feist, M. I., & Gentner, D. (2003). Factors Involved in the User of In and On. *Proceedings of the Twenty-fifth Annual Meeting of the Cognitive Science Society*.
- Gapp, K. P. (1995). *Object localization: selection of optimal reference objects*. Springer Berlin: Heidelberg.
- Kelleher, J., & Kruijff, G. J. (2006). Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of the 21st International Conference on Computational Linguistics (COLING) and 44th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp.1041–1048), Sydney, Australia.
- Kelleher, J., Ross, R., Mac Namee, B., Sloan, C. (2010). Situating Spatial Templates for Human-Robot Interaction, in Dan Bohus, Eric Horvitz, Takayuki Kanda, Bilge Mutlu, Antoine Raux (eds.) *AAAI Fall Symposium: Dialog with Robots*, California: AAAI Press.
- Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73, 31–68.
- Krahmer, E., & Theune, M., (2002). Efficient context-sensitive generation of descriptions in context. In K. van Deemter & R. Kibble, (Eds), *Information Sharing: Givenness and Newness in Language Processing*. Stanford: CSLI Publications,
- Krahmer, E., & van Deemter, K. (2012). Computational Generation of Referring Expressions: A Survey. *Computational Linguistics*, 38, 173-218.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, London: The MIT Press.
- Maas, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: nature or culture? *Psychological science*, 14, 296–301.
- Mandler, J. M. (1992). How to build a baby: II Conceptual primitives. *Psychological review*, 99, 587-604.
- Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Miller, J. E., Carlson, L. A., & Hill, P. L. (2011). Selecting a Reference Object. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 4, 840 – 850.
- Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and

- Spanish. *Journal of Memory and Language*, 42, 168-182.
- Talmy, L. (1983). How Language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial orientation: Theory, research and application* (pp. 225–282). New York: Plenum Press.
- Tversky, B. (2011). Visualizing Thought. *Topics in Cognitive Science*, 3, 499–535.
- Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective taking. *Cognition*, 110, 124-129.
- Tversky, B., Kugelmass, S. & Winter, A. (1991). Cross-cultural and developmental-trends in graphic productions. *Cognitive Psychology*, 23, 515-557.
- van der Zee & Watson, M. (2005). Between Space and function: How spatial and functional features determine the comprehension of between. In L. Carlson & E. van der Zee (Eds.) *Functional Features in Language and Space*, New York: Oxford University Press.
- Viethen, J., & Dale, R. (2008). The use of spatial relations in referring expressions. In White, M., Nakatsu, C., McDonald, D. (eds.) *Proceedings of the 5th International Conference on Natural Language Generation*, (pp. 59-67). USA: Salt Fork OH.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Zipf, G.K., (1949). *Human Behavior and the Principle of Least Effort: an Introduction To Human Ecology*, Addison-Wesley: Cambridge